知识管理论坛 ISSN 2095-5472 CN11-6036/C
Knowledge Management Forum
E-mail: kmf@mail.las.ac.cn
http://www.kmf.ac.cn

【学术探索】

基于知识图谱的数字档案服务模式探究

熊回香 严舞月
华中师范大学信息管理学院 武汉 430079

摘要：[目的/意义]针对当前数字档案服务质量智能化程度不足、服务内容单一等短板，设想构建数字档案知识图谱整体架构，达到数字档案大数据的统计分析、数字档案资源集成优化以及数字档案整体服务水平提升的目的。[方法/过程]通过收集文献分析数字档案馆的服务缺陷，体验式调研各省市数字档案馆网页服务水平，汇总各档案馆目前服务模式的不足，完善数字档案知识图谱架构流程，最终以流程图的形式展示。[结果/结论]知识图谱能够将多种类型的数字档案文本转化为计算机可理解的数据，提高计算机智能识别水平，同时图谱所具备的动态时序性和针对性能够按照时间节点提高档案整合程度，同时基于用户浏览检索痕迹更新用户数据，提升服务质量，并增强数字档案馆之间的合作交流以达到整合资源的效果，为更好地优化数字档案服务提供有益参考。

关键词：知识图谱；数字档案；档案智能服务
分类号：G270

引用格式：熊回香，严舞月．基于知识图谱的数字档案服务模式探究[J/OL]. 知识管理论坛，2021, 6(4): 204-212[引用日期]. http://www.kmf.ac.cn/p/254/.

1 引言

由于语义技术的突出表现，许多领域都看到了语义网络带来的便利及其难以被取缔的优势，档案界也开始对数字档案的未来发展有了新的想法。同时随着互联网的发展，人与人、档案与人的交流也愈发密切，档案服务与语义网络的结合已经在所难免。目前数字档案资源的组织在语义方面尚处于初步尝试阶段，如何将数字档案资源在语义层面组织起来，为用户提供更为精准的服务，已成为当前档案服务部门亟待解决的现实问题。基于当前数字档案服务所存在的零散、复杂和智能化程度不够等问题，本文提出将知识图谱技术与数字档案服务融合这一理念。知识图谱作为一种新兴语义处理模型，能够将实体与实体连接起来，挖掘并

基金项目：本文系国家社会科学基金年度项目“融合知识图谱和深度学习的在线学术资源挖掘与推荐研究（项目编号：19BTQ005）研究成果之一。
作者简介：熊回香（ORCID：0000-0001-9956-3396），教授，博士，博士生导师；严舞月（ORCID：0000-0003-1652-7926），硕士研究生，通讯作者，E-mail: 348101761@qq.com。
收稿日期：2021-05-13 发表日期：2021-08-19 本文责任编辑：刘远颖

ChinaXiv合作期刊
知识管理论坛，2021（4）:204-212
DOI: 10.13266/j.issn.2095-5472.2021.020

展示实体间关系，提炼不同档案核心词，将档案相关知识进行整合，促进数字档案资源聚拢，提升用户使用体验，达到用户输入单一档案知识点，便能通过知识图谱来获得其他相关档案知识的效果，即图谱自动实现相关内容扩展，减少检索步骤，提高检索效率，完善检索内容，最终实现优化智能服务的目的。

2 相关研究

知识图谱是一种描绘实体之间关系的语义网络，是人工智能重要研究领域——知识工程的主要表现形式之一。目前可获取的相关知识库资源包括国外的 Freebase、Wikidata、DBpedia、YAGO 等，国内有复旦大学公布的中文概念图谱 CN-Probase 等。本文构建的知识图谱架构不是泛化的通用知识图谱架构，而是构建基于档案的领域知识图谱架构。不同于通用知识图谱，领域知识图谱能利用领域特有知识快速构建知识库，如医疗知识图谱、地理知识图谱、军事知识图谱及农业知识图谱等[1]。目前，知识图谱的研究主要集中在针对已有元数据（EAD、Dublin Core 等）的基础上，探讨元数据语义互操作以及映射关系[2]。例如，杨茜雅在企业档案数据应用中引入语义本体概念实现档案数据语义分析的流程，在此基础上构建联通电子档案知识图谱系统[3]；雷洁等基于 Protégé、OWL 等技术构建计算机可理解的科研档案知识图谱语义模型[4]；舒忠梅基于当下数字人文的发展背景，提出档案时空本体模型及档案数据抽取框架，构建档案关联数据知识图谱，以可视化的形式展现[5]；B. S. Balaji 等采用语义对描述文档进行解析，并构架云服务推荐系统[6]。

而在数字档案服务方面，以往用户熟知的档案服务方式多为被动服务，即被动调动和被动查看，导致各档案文件之间多呈现孤立关系，用户取用困难，操作繁琐。目前国内大部分学者使用转变服务模式、构建资源平台等方法来解决该问题。例如，曹玲等对美国常青藤八所高校的数字档案信息服务模式进行研究，将其分为运营方式、服务对象、服务方式三个模块进行分析，提出基于我国档案服务优化建议[7]；王文强通过分析智慧服务和数字档案馆的利弊，转变企业档案信息服务模式，由“信息服务”向“知识服务”“智慧服务”模式优化[8]；连志英基于对数字档案信息用户需求、用户行为的分析，对数字档案信息用户进行分类及构建数字档案信息用户模型，并且根据用户需求和用户行为建设数字档案信息资源及数字档案信息服务平台，用以提供相应的个性化数字档案信息服务[9]。国外，大多是将数字档案与文化遗产保护以及历史应用相结合，也会相应探讨一些关于数字档案馆建设的内容。例如，T. Hauswedell 等考察了制度、知识、经济、技术、实践和社会因素的复杂相互作用，与主要报纸数字化计划的公营和私营供应商进行了一系列半结构化的采访并进行了分析，认为那些很少被突出或强调的因素的新兴理解，从根本上塑造了数字文化遗产档案的深度和范围，应关注这些因素的在未来档案发展中的潜在优点[10]；C. H. Marcondes 分析了档案馆等使用关联数据技术的可行性及问题[11]。

综合以上所述，目前针对数字档案的服务大体上还处于一种针对服务内容和服务框架的构建，集中在对资源、数据、服务模式等的探讨方面，实践性研究较少。且能明显感觉到当下各省市数字档案的系统性管理和智能服务提供并不完备，目前国家正快速迈进智慧时代，随着对语义网络相关技术的逐步了解，学者也充分意识到语义技术的发展对于数字档案服务效率以及服务质量的提升有着重要的影响。而知识图谱所带来的集成化和系统化可以很好地应对数字档案零散化问题，并且通过整理推荐为用户提供有针对性、全面性的服务。另外，基于数字档案服务在各行各业的不同作用情况，本文考虑数字档案的统一特征，通过语义关联，联结数字档案实体，运用可视化图表展示实体间关系，构建知识图谱以推进数字档案服务模式的升级。

3 数字档案服务模式现状分析

3.1 现有服务模式的局限

档案服务是指档案机构利用馆藏优势，指导用户利用档案、获取档案信息的过程[12]。

而现有衡量服务模式优劣的两个重点包括档案服务人员服务质量以及档案服务机构智能化程度，在档案服务人员服务质量方面，传统的档案服务模式多为被动服务，大多数企业或政府的档案工作人员在档案服务整体信息化程度不高的情况下，对档案服务的认知不够明确，其职能仅仅在于管理和保存纸质档案，维护档案信息安全等，而为用户提供档案检索服务等更具有实际意义的工作不在其职责范围内，将如何从数量庞大的档案信息数据库中查找所需档案的难题留给了用户自己。

另外，在档案服务机构智能化程度方面，由于新时代“互联网＋大数据”的飞速发展，档案服务走向主动和智能的模式，通过计算机设备与其他各种互联网终端相连接，通过一定的技术方式向用户提供各类档案信息或产品的服务模式，逐渐成为数字档案服务模式的主流，其中最具代表性的是数字档案馆和智慧档案馆。但目前数字档案馆的建设还处于起步阶段，所提供的个性化服务极其有限，不仅智能化程度低，且操作复杂，难以形成整合型的服务进行推送。笔者对现有省市档案馆网站进行体验式调查后总结发现，现有档案馆网站中虽然大多拥有开放档案查询服务，但基本依托用户自主查询，且对查询关键词要求较高，同时跳转链接较多，查询过程复杂，例如，登录湖北省档案馆官网查询某档案，系统提示进入档案信息网进行查询，但并没有提供对应链接。通过自行查找得到的档案信息网链接，点击得出结果则直接跳转回湖北省档案馆主页。因此可以看出，当前档案馆网站存在服务水平有限、集成化程度不高、操作复杂等局限性。

3.2 基于知识图谱的数字档案服务模式的可行性分析

随着语义网络、神经网络的发展，知识图谱、图数据库、深度学习等相关技术也逐渐渗透到图情档研究领域，使得档案数据语义描述与关联、知识融合、信息可视化等成为可能。同时，伴随智慧城市的建设推动，档案服务也开始重视用户需求并将服务方式逐渐向以用户需求为中心转移，诸如数字档案馆、智慧档案馆等，在处理档案数据、调动档案资源等方面尚存不足。基于知识图谱的数字档案服务通过对数字档案资源的语义关联和本体提炼，可以准确提取数字档案资源的核心，并通过简洁明了的可视化形式展现给用户。

首先，档案数据质量较高。相较于数量庞大且形式复杂的互联网大数据，档案数据都是真实事件记录和数据保存，在入库时已经经过筛选、分类和整理，对不同的数字档案也有相应的规范格式和要求，这就为基础档案数据处理构建了良好的基础；其次，知识图谱的系统性能够有效帮助整理零散的档案。即运用图映射、包装器等工具，基于对数据深加工的需求，对数字档案的结构进行清洗、变换和集成，使之变为计算机可以理解的结点，在此基础上识别实体、连接实体、分析实体，对档案实体进行语义关联，连接成数字档案知识图谱。最后，知识图谱的输出模式能够有效提升用户体验。图谱将用户搜索到的相关档案以可视化的形式输出，能够更加清晰和便利地展现档案间的联系，同时图谱的动态更新特性能够及时修改用户取向，使得服务更加智能。

4 基于知识图谱的数字档案服务模式架构

知识图谱的构建通常有自顶向下和自底向上两种模式[13]，基于档案的强领域性，本文选择自底向上的构建形式，即通过迭代更新，从信息抽取到知识融合、知识加工，最后进行知识更新，从分析数据到输出服务的模式。基于知识图谱的数字档案服务模式最终回归于应用领域，将知识图谱与数字档案服务模式相结合的目的也是为了优化服务模式，因此依据知识图谱理论、本体理论、智能代理技术、云计算

知识管理论坛，2021（4）:204-212
DOI: 10.13266/j.issn.2095-5472.2021.020

技术以及个性化推荐技术，将数字档案知识图谱服务平台梳理为 4 个层次：①知识抽取层。主要作用是将外部不同结构的数字档案信息资源汇集起来，通过知识抽取层的结构化分类和分层，转化成计算机可识别和理解的数据，规整数据使其达到标准化格式从而进入图谱构建层。②图谱构建层。这一层主要进行档案知识抽取和本体构建，然后通过实体识别和本体对齐形成数字档案资源知识库。③图谱更新层。该层主要分为两个部分，包括收集用户浏览痕迹对用户个体数据库进行更新，以及收归新档案对数字档案资源库进行更新。通过这两方面的更新使得构建的数字档案知识图谱不是单一的、片面的，而是不断完善的。④图谱应用层。最终构建的数字档案知识图谱能够为档案服务带来不同的应用效果，包括图谱动态更新、用户偏好个性化推荐、数字档案智能搜索和数字档案馆资源共享。具体架构如图 1 所示：

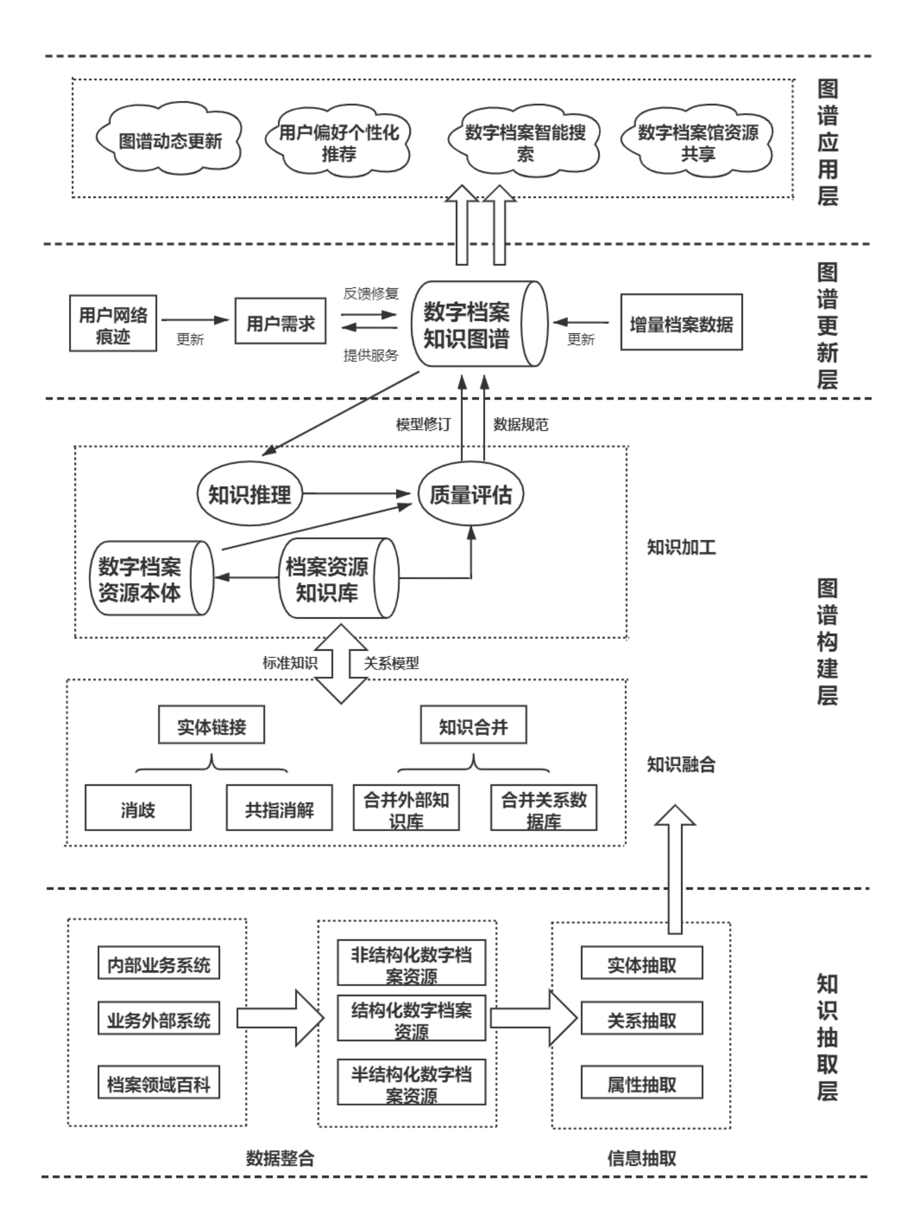


图 1 基于知识图谱的数字档案服务模式架构

4.1 知识抽取层

知识抽取层主要包括数据整合和信息抽取两个部分。数据整合部分是对收集到的不同数字档案数据进行整理和挑选，这些数据包括从不同行业内部业务系统中收集的、从档案领域百科中整理出来的以及从业务外部系统承接的，不同形式和不同结构的数字档案数据。而档案数据包括各级各类档案机构收集的具有档案性质的数据记录，包括各种数据形式的档案资源，如各类数字档案、多媒体档案；亦包括档案管理与利用过程中产生的数据，如档案网站的浏览记录、平台日记、查阅服务数据、档案统计数据等 [14]，而日常办公使用的 Word、PPT、Excel、PDF 等数据是非结构化数字档案资源 [15]。基于当下数字档案资源的结构化不统一的问题，笔者认为可以运用自然语言处理技术（NLP）以及包装器等工具将这些非结构化资源进行统一调整、清洗和修正，包括运用词典、统计和规则的方法对档案数据进行分词，再基于分词结果进行清洗过滤，如去掉停用词、去除单字，对分词进行词性标注等。同时还可以利用 Word2vec 等模型通过词嵌入将档案词语从 one-hot encoder 形式的表示降维到较短的词向量，使得计算机能够更好地理解和分析数据，发现实体之间的语义关系。而信息抽取部分则是包括实体抽取、关系抽取和属性抽取三个部分，其中，档案实体抽取是运用规则与词典、统计机器学习和面向开放域三种方式，从数字档案资源中识别并提取实体；档案关系抽取是指采取监督学习或远程监督学习的方式，抽取实体间的关系，解决档案资源实体间语义链接的问题；档案属性抽取是指对档案资源实体的某些特征和性质进行抽取，也可以看作是实体与属性值间的一种名词性关系，具体流程如图 2 所示：

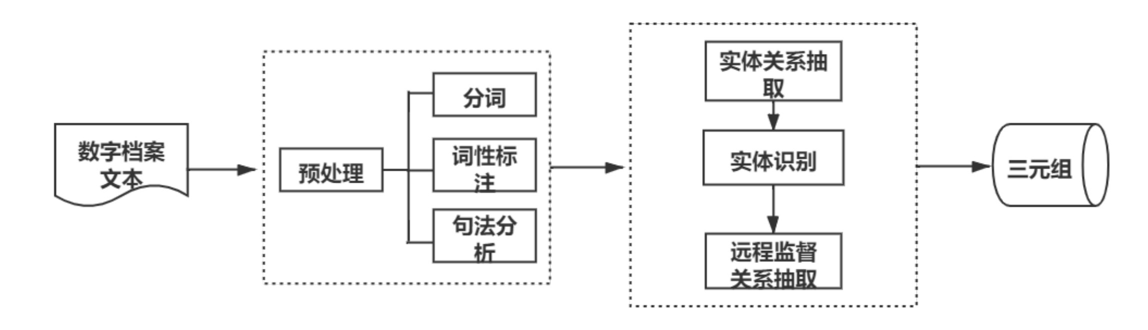


图 2 数字档案领域实体抽取与关系抽取

4.2 图谱构建层

图谱构建层分为知识融合和知识加工两大部分。在大量非结构化档案信息经过处理后，能够获得实体、关系以及属性的相关信息，但这些结果可能包含大量的错误信息和冗余重复信息，数据之间的关系也不清晰，缺乏层次性和逻辑性，因此需要通过知识融合进行清洗和整合。知识融合包括两个方面，其中实体链接是指将抽取到的实体与知识库相联结的程序，主要方法包括实体消歧和共指消解。实体消歧主要是解决同名实体出现歧义的问题，使用较多的方法主要为聚类法。而共指消解则是用于解决多个指向对应一个实体对象的问题，国外相关研究相对来说已经比较成熟，同时出现的实体相似性模型、上下文相似性模型能够很好地解决这一问题。而知识合并是指将第三方的数据合并进数据库，包括合并外部知识库和关系数据库两个方面。

通过知识融合，可以得到一系列基本的事实表达，但事实并不等于知识，因此需要进入知识加工阶段。知识加工主要包括三方面内容：本体构建、知识推理和质量评估。数据在经过

知识管理论坛，2021（4）:204-212
DOI: 10.13266/j.issn.2095-5472.2021.020

知识融合之后，识别实体已经变成标准化知识并且附有相关属性关系，被归入知识库中。受现有技术限制，通过信息抽取的知识元素仍可能存在错误，因此在构建完整的档案知识图谱之前，需要进行质量评估，并且通过对知识的可信度进行量化，通过舍弃置信度低的知识来确保档案知识图谱内数据的质量。同时从已有的实体关系出发，经过计算机推理，发现新的实体关系，档案知识图谱网络得到进一步完善和更新。而数字档案资源本体，是经过组织的一种质量较高的知识表示模型，在知识图谱的构建中本体主要可起到控制图谱质量的作用，不一定会参与到图谱建立的流程中，仅作为一种数据质量评估的参考资源库，在档案知识图谱构建过程中充当辅助角色，同时帮助优化知识图谱更新。

4.3 图谱更新层

图谱更新层包括用户需求更新和档案资源更新两大模块，通过收集归纳新进档案和用户网页浏览查询痕迹等数据，保持图谱的实时性，同时提高服务精准度，增强图谱适配性。

在用户需求更新方面，不同的用户行为数据通过数据驱动自动对档案资源进行本体构建，再经过质量评估方法与人工审核相结合的方法加以修正与确认。本体构建之后，对知识库的数据来往以及反馈不断进行修订，同时，知识库也收集用户对知识库的使用痕迹并不断进行调整和更新，加入时间维度，利用时序分析技术和图相似性技术，分析图谱结构随时间的变化和趋势，从而掌握到关键信息，构建动态时序图谱。

在档案资源更新方面，运用知识图谱中的知识推理板块，如基于 Tableaux 运算能够检查某一本体的可满足性，同时通过实例对本体进行检测；而基于逻辑编程改写可以根据特定的场景定制规则，以实现用户自定义的推理过程；基于一阶查询重写能够高效地结合不同数据格式的数据源，同时关联起不同的查询语言；基于产生式规则可以控制系统的执行，通过制定一定的机制执行规则实现更好地前向推理等；另外，针对构建的知识库进行质量评估也是确保知识图谱内容正确可用的关键步骤，评估结束后需将符合标准的档案资源数据导入档案知识图谱中。

4.4 图谱应用层

图谱应用层是指通过完整的知识图谱构建流程，将数字档案馆中不同的数字档案资源进行集成和整理，形成档案知识图谱，输出到应用层面，通过知识图谱对数字档案信息源的生成数据进行处理，将产出的结构化关联数据用于深度学习算法训练，得到能解决具体场景问题的研判模型，从而形成解决办法产生价值的服务形式，包括基于关联规则算法的图谱动态更新、基于聚类算法的档案用户偏好个性化推荐、基于分类与预测算法的数字档案智能搜索、基于整体优化的数字档案资源共享等。

5 基于知识图谱的数字档案服务模式优势及应用

基于知识图谱的数字档案服务模式架构的构建始终立足于为用户提供更有效、更方便、更智能的服务，通过对数字档案的数据整理，结合知识图谱构建流程，利用自然语言处理技术、实体识别、本体构建、关系抽取等关键技术，构建基于数字档案知识图谱的架构，能够支撑数字档案的智能性管理，提供动态档案图谱智能更新、自动分析用户偏好、立足数字档案内容的智能搜索以及数字档案馆资源共享。

5.1 图谱动态更新

随着互联网技术的不断发展，档案服务逐渐实现数字化，但目前的数字档案服务现状仍浮于表面，首先建设的是数字档案的存储与管理问题，对于优化数字档案服务的内容没有较多的探讨。基于数字档案的知识图谱的构建，可以通过数字档案本体之间的关联关系，使得档案内容被更加方便快捷地分解和分类，再通过图谱关系梳理，使得档案相关内容能够产生联结，在用户检索其一时将相关内容完整地推

送出来。数字档案管理不同于数字图书的管理，由于档案具有隐私性，因此新的数字档案归档时需进行隐私性和公开性的衡量，从而导致在档案入库时程序更加繁琐和复杂。基于语义关联的知识图谱的构建，可以使数字档案文本入库之时就通过识别本体创建链接，自动分类。而动态更新则是指图谱的构建不是一成不变的，而是随着档案的增加和删减而不断变化和完善，可以通过加入时间、空间等维度，构建时间轴或空间轴来完善图谱在不同时空的内容。图谱的动态更新能够有效提高档案整理效率，同时便于不同类型档案的规整，也能够为档案用户提供最新内容。当前工程档案是数字档案收藏和管理的重点，如港珠澳大桥这类国家大型工程，其档案拥有耗时长、总量大、部门多等特点，知识图谱的动态更新特性能够很好地跟上工程档案归档需求，全程记录归纳总结，自动分类整理，提供更加便利的服务。

5.2 用户偏好个性化推荐

《全国档案事业发展“十三五”规划纲要》明确指出，“要提高档案公共服务能力，提升档案服务的认知度和用户满意度”。档案用户面对当前繁杂无序的档案资源，不仅需要自己提炼检索语句，还需要处理复杂的检索程序，导致增大检索耗时，也会降低用户检索兴趣，消减用户检索需求。同时，在通过图谱提供个性化服务时，档案用户的需求会随着用户行为和浏览的档案内容而产生实时变化，用户需求的易变性、多样性与数字档案内容和类型多样性的关联，提高了档案个性化服务功能预测用户需求的难度。因此，针对用户的动态行为反向推断用户档案偏好对于提升数字档案服务十分重要。基于前期知识加工对档案数据的转化，使用自然语言处理和机器学习方法对档案资源中的实体进行概念提取、类及其等级体系的确定、类的对象属性及数据属性的确定，以及本体评价等过程，选取 BERT 和 LSTM 等多类深度学习模型 [13]。预测系统不仅能够根据用户页面停留时间以及浏览速度来衡量用户偏好，并且能够运用相关技术来识别用户的自然语言，通过标准化处理转化为计算机可识别语言，让计算机能够更好地识别用户偏好，增加数字档案服务的准确性。在预测用户需求的同时，可以根据已收集到的用户行为来为下次服务做铺垫。例如，用户对实时新闻以及档案趣事感兴趣，知识图谱识别到关键词后根据语义关联会相应地推送实时趣事，以及与趣事产生联系的一系列相关内容，若用户对历史档案以及领域专家研究感兴趣，则会推送相应学术资源等。

5.3 数字档案智能搜索

档案用户使用传统档案检索时，只有输入准确的档案关键词搜索题名或内容，才能检索到相关信息。这种服务模式对检索语言精准性要求较高，甚至有时出现由于档案用户无法准确表达关键词而搜索不到想要的档案资源的情况。而基于知识图谱的数字档案服务模式，通过自然语言处理技术（NLP）能够将用户搜索时使用的自然语言自动转化为计算机可以理解的语言，同时映射到数字档案知识图谱中不同的实体或属性层级，通过结合实体间的关系来推送相应的数字档案信息资源，提供体系化的档案知识供用户浏览。用户使用自然语言进行检索时也能精准反馈相应内容，提高档案查询的查准率和查全率，并且提供内容联想服务，将相关内容统一且连贯地展现在用户面前，减少用户查询次数和查询步骤，提高检索效率。“档案潜在需求是未被唤醒或未被认识到的需求，主要包括档案潜在用户的需求及档案现实用户未表达出来的真实的需求” [16]，这种智能内容搜索能够帮助数字档案馆加深与用户之间的联系，通过相关内容集成推送的创新服务，吸引用户浏览，同时帮助档案馆收获更多潜在需求被挖掘的用户，提升服务质量。

5.4 数字档案馆资源共享

基于知识图谱的数字档案服务，将不同数字档案馆的资源聚集，通过云存储、云计算等技术整合在一起，建立数字档案知识图谱，促进资源间的相互交流，不仅便捷地服务用户，

知识管理论坛，2021（4）:204-212
DOI: 10.13266/j.issn.2095-5472.2021.020

也给各大档案馆之间的交流架起桥梁。为了迎合当下信息资源相互沟通的大环境，档案界资源交互也显得尤为重要。基于知识图谱的数字档案服务模式，能够将不同数字档案馆中的资源汇集，通过识别本体来进行语义联结，再基于档案领域关键词将其自动归类于不同行业，便于查找利用。知识图谱模型的帮助使得不同数字档案馆之间的资源共享具有很强的可操作性。在服务用户方面，资源的融合能够带来更丰富的档案内容，达到减少检索步骤的同时丰富检索内容的目的。基于当下互联网带来的便利，很多信息都可以如实地通过互联网检索到，但由于档案所具备的隐私性和个体性使得档案检索受限于地区资源。知识图谱的资源共享所带来的便利能够很好地解决地区间的信息孤岛，并且通过融合加深档案馆之间的创新合作，共同提升服务质量，让档案用户真正了解档案世界中丰富的内核，扩大档案受众群，使得民众提高档案利用意识，提升档案利用价值。同时，数字档案馆在开展资源共享服务时也可依据定位需要选择不同的机构来建立战略联盟。在选择合作伙伴时，可选择同一领域特长的数字档案机构，从而加强某一领域档案的馆藏量，体现自身在某一领域的特色；也可以选择不同领域特长的档案机构，从而提高自身的综合性[17]。通过与不同档案馆的合作，结合知识图谱所带来的沟通性和联结性，能够更好地促进数字档案服务机构的资源交互，为数字档案服务提供新的交流环境，推动数字档案服务升级。

6 结语

信息化时代，人们的信息行为、方式等各个方面都经历着前所未有的改变。语义网、大数据及深度学习等技术的快速发展，为数字档案的准确性和智能性服务提供了技术保障。本文提出一种基于知识图谱的数字档案服务模式，充分运用符合当代互联网技术发展潮流的新型智慧技术来解决当下档案服务所面临的问题，针对当前数字档案服务中存在的检索语言难识别、检索步骤繁杂、检索内容单一等问题，通过创建知识图谱架构，提出解决策略，不断推进档案服务发展以及档案业务与新兴技术的结合。但本研究仅针对数字档案服务进行探讨，仍存在浮于理论、难以实现等问题，未来将努力推进数字档案知识图谱的现实构建，并通过收集真实用户反馈意见对该构想进行进一步完善，促使数字档案服务有更深层的提升和优化。

参考文献：

[1] 王电化 , 钱涛 , 钱立新 , 等 . 面向档案的知识图谱构建方法研究 [J]. 湖北科技学院学报 , 2020, 40(1):127-130.

[2] 雷洁 , 李思经 , 赵瑞雪 , 等 . 面向科研档案管理的知识图谱构建与应用研究 [J]. 数字图书馆论坛 , 2020(5):8-15.

[3] 杨茜雅 . 中国联通电子档案数据挖掘与智能利用的研究 [J]. 档案学研究 , 2018(6):105-109.

[4] 雷洁 , 赵瑞雪 , 李思经 , 等 . 知识图谱驱动的科研档案大数据管理系统构建研究 [J]. 数字图书馆论坛 , 2020(2):19-27.

[5] 舒忠梅 . 数字人文背景下的档案知识图谱构建研究 [J]. 山西档案 , 2020(2):53-60.

[6] BALAJI B S, KARTHIKEYAN N K, KUMAR R. Fuzzy service conceptual ontology system for cloud service recommendation[J]. Computers & electrical engineering, 2018(69):435-446.

[7] 曹玲 , 王榕 , 颜祥林 . 分析与借鉴美国常青藤高校数字档案信息服务模式 [J]. 数字与缩微影像 , 2013(3):30-34.

[8] 王文强 . 基于数字档案馆的企业档案智慧服务模式探析 [J]. 机电兵船档案 , 2019(4):76-78.

[9] 连志英 . 基于用户需求的个性化数字档案信息服务模式构建 [J]. 档案学通讯 , 2013(5):49-53.

[10] HAUSWEDELL T, NYHAN J, BEALS M H, et al. Of global reach yet of situated contexts: an examination of the implicit and explicit selection criteria that shape digital archives of historical newspapers[J]. Archival science, 2020, 20(2):139-165.

[11] MARCONDES C H. Interoperability between digital collections in archives, libraries and museums: potentialities of linked open data technologies[J]. Ciência da informação, 2016, 21(2): 61-83.

[12] 张卫东 , 王萍 . 档案用户需求驱动的个性化服务模式研究 [J]. 档案学通讯 , 2007(2):82-86.

[13] 刘峤，李杨，段宏，等．知识图谱构建技术综述 [J]. 计算机研究与发展，2016, 53(3):582-600.
[14] 赵跃．大数据时代档案数据化的前景展望：意义与困境 [J]. 档案学研究，2019(5):52-60.
[15] 李超．视频侦查的知识图谱构建研究 [D]. 北京：中国人民公安大学，2019(6):1-25.
[16] 杨静．档案潜在用户研究 [D]. 合肥：安徽大学，2013:9.
[17] 赵宏育．如何做好档案机构间的档案交流 [J]. 兰台世界，2020(S1):12.

作者贡献说明：
熊回香：研究内容指导；
严舞月：论文撰写与修改。

Research on Digital Archives Service Mode Based on Knowledge Graph

Xiong Huixiang Yan Wuyue
School of Information Management, Central China Normal University, Wuhan 430079

Abstract: [Purpose/significance] Aiming at shortcomings of the current service quality of digital archives, such as insufficient intelligence and single service content, this paper proposed to build an overall framework of knowledge graph of digital archives, so as to achieve statistical analysis of big data for digital archive, integration and optimization of digital archiving resources, and improvement on the overall service level of digital archives. **[Method/process]** By collecting literatures, service deficiencies of digital archives were analyzed. This paper investigated the web page service level of digital archives in various provinces and cities, summarized deficiencies of the current service mode of each archive, and improved the framework process of digital archives' knowledge graph, finally the framework process was presented in the form of flow chart. **[Result/conclusion]** Knowledge graph can transform multiple types of digital archive text into data that computers can understand, and improve the level of computer intelligent identification. At the same time, dynamic timing and pertinence of the graph can improve the degree of integration archives according to the time node, user data can be updated based on users' retrieval trace to improve the service quality, enhance the cooperation and communication between digital archives to achieve the result of resources integration, and provide a useful reference to better optimize the digital archiving service.

Keywords: knowledge graph digital archives archive intelligence service